

REAL-ACT 01 | China-US AI incident communication channel

Complete Web Text PDF Edition

Exported from the public website edition on 2026-10-17. Includes the complete article, edition and evidence boundaries, references, and further reading. This is a separate downloadable web-text edition, not another DOCX master or a verified original publication. Statements below about unattached or unproduced PDFs reproduce the website at export; they do not refer to this new web-text export.

The webpage includes text transcribed from source tables. This edition retains that content and reflows line breaks; it does not reconstruct the original table layout. Facts and evidence levels are the source text's claims at its stated cutoff, not independent verification of every external event during this export.

<https://www.civitas.top/en/documents/p03-surv-004/>

Verified public source and edition

Reference [2] has been located at the Ministry of Foreign Affairs alternate official host. The 26 September 2026 article and point seven support the cited agreement on an AI dialogue, a November meeting and an AI incident communication channel. This verifies the cited political commitments, not implementation or effectiveness; the source evidence cutoff remains 28 September 2026.

Verified official reference [2]

https://www.fmprc.gov.cn/zyxw/202609/t20260926_12031616.shtml

Translation scope

This page translates the complete publicly available Chinese HTML text, including its source version and evidence boundaries. It does not claim that a separate English PDF has been produced. Source transcription and original-document formatting limitations remain as stated below.

The problem this work addresses

Place protocols, time, changing power and the regeneration of correction within a real high-consequence window. Strictly distinguish political commitments, operational rules, actual use and effects.

Version and evidence boundaries

Current version: V0.2.1. This English page translates the public online transcription of the uploaded Chinese source document. The source PDF has not yet been attached through this site's binary publication process, so no invented download link is provided. Refer to the source document for tables and complex formatting.

Continue reading

- Return to the underlying logic: why protocols are not stability itself

<https://www.civitas.top/en/documents/p03-surv-002/>

- See the wider network through a real event

<https://www.civitas.top/en/documents/p03-global-001/>

- Continue to states' long-term interests and shared survival

<https://www.civitas.top/en/documents/p03-only-002/>

Complete online text

Civilization Leap Research Group · REAL-ACT 01 · A Real-World Problem Tested Across Time

REAL-ACT 01

A China-US AI Incident Communication Channel

—Agreements constrain existing strategic games; self-interest first keeps generating new ones: from provisional agreements to the capacity to regenerate correction

V0.2.1 | Edition calibrated for the foundational judgment and further reading

Evidence cutoff: 2026-09-28 | Author: Zijun Fu | Produced by the Civilization Leap Research Group

Agreements can constrain strategic games already under way, but cannot prevent a self-interest-first logic from continually generating new ones. What civilization must carry through time is not an eternal agreement, but the capacity to regenerate correction and peace before irreversibility whenever old agreements, actors, and equilibria cease to hold.

This text follows the real-world window identified in REAL-ACT 01 V0.1: China and the United States have publicly agreed to establish AI/SI dialogue and incident communication channels. V0.2 no longer makes whether a channel can be established its main question. It extends the time horizon: why would this structure still exist after a change of governing party, a major error of judgment, a social transformation, a technological leap, or a reversal of power? If it cannot endure such changes, today's agreement is only a provisional equilibrium. If it can, what deserves study is not the agreement's text itself, but the conditions that continually regenerate the capacity for correction.

I. The real-world window still exists, but is not yet evidence of long-term stability

As of September 28, 2026, public materials support the following facts:

- White House materials state that China and the United States established a Super Intelligence (SI) Dialogue to exchange views on associated risks and benefits, that the next exchange would occur by November 2026, and that they agreed to establish a bilateral communication channel for SI incidents.[1]
- The Chinese Ministry of Foreign Affairs' published "eight-point consensus on outcomes" states that the parties agreed to establish a China-US AI dialogue, hold the next dialogue in November 2026, and establish an AI incident communication channel. It also states that the two militaries agreed to sign a memorandum of understanding on strengthening crisis communication and prevention as soon as possible.[2]
- China's account of the leaders' talks proposes continuing dialogue on AI risks and benefits and jointly preventing AI misuse, emphasizing that AI should remain under human control.[3]
- Public materials have not disclosed complete triggers, disclosure scope, response deadlines, verification procedures, false-alarm handling, attribution of responsibility, or rules separating this channel from other issues.

At this stage, therefore, the most that can be confirmed is the emergence of political commitments and an entry into a mechanism. This does not establish that the channel is operating or has reduced risk, still less that it can endure changes in governing parties, power, and technology.

II. Why an agreement is not stability itself: reasonable at signing does not mean future actors still have reasons to comply

An agreement usually emerges from a particular configuration of actors, interests, power, information, and risk. When those conditions change, the judgment that compliance is more advantageous than violation may change too. International-relations research does not support an absolute conclusion that agreements must fail, but it does show that changes in power and decision processes affect the reliability of commitments.

- Fearon's classic study identifies the inability to credibly commit to continuing mutually beneficial arrangements as an important mechanism in rationalist explanations for war. Even when both sides prefer a peaceful arrangement today, it may fail because future commitments lack credibility.[4]
- Leeds's study of alliance commitments from 1816 to 1944 finds that most commitments in its sample were honored, but post-signing changes in national power or policymaking processes were important predictors of violation.[5]
- Powell further formalizes commitment problems under rapid changes in power: an arrangement acceptable today may become noncredible because the future distribution of power changes.[6]

These studies do not prove for this essay that all agreements eventually fail. They show that agreements' stability must confront time: today's equilibrium cannot automatically be extrapolated into the future.

The hardest terrain for a peace agreement is not today, but the people, powers, and eras that have not yet appeared.

III. The real stress test: place the agreement within the long course of time

REAL-ACT 01 no longer asks only whether the channel will be more fully established in the next dialogue. It asks whether it can withstand at least six kinds of change.

Temporal shock | The real question | What failure exposes beyond a particular written clause

T1: Change of governing party/government | Why would new decision-makers continue honoring old commitments? | Commitments depend only on the previous actors' intentions; stable reasons across actors have not formed.

T2: Major error of judgment | Would a leadership misjudgment, faulty intelligence, or mistaken attribution close communication first? | The security structure still makes civilization's insurance depend on key individuals always judging correctly.

T3: Social transformation and reconstitution of actors | Who inherits commitments when the former "we" splits, reorganizes, or encounters new interest-bearing actors? | The agreement binds old actors rather than a logic of correction that new actors can regenerate.

T4: Reversal of power | Why refrain when one side first discovers that breaking rules could confer a clear advantage? | The agreement loses its incentive foundation precisely when it is needed most.

T5: Technological leap | Can the agreement reopen after new AI/AGI capabilities alter the meanings of incident, attack, autonomous behavior, and responsibility? | Clauses are fixed while the objects they address have changed.

T6: Crisis and failure | After false alarms, leaks, exploitation, or a severe failure, is the mechanism abandoned or can it be rebuilt? | Correction of the correction mechanism itself has not been preserved.

These six changes need not occur within the same country, and they do not presume any particular political system is necessarily more or less stable. They show only that long-term mechanisms must face changing actors, power, understanding, and technology.

IV. What must endure through time is not the agreement's text, but the capacity to regenerate correction

It is unreasonable to define stability as a 2026 agreement remaining word for word unchanged a century later. Future actors, technologies, and risk structures may be entirely different. Civilization more urgently needs to preserve a capacity: to generate new agreements before old ones fail; to let new actors rebuild shared security after old actors change; to form new equilibria after one side grows stronger without first passing through irreversible conflict; and to keep channels of correction alive after a leadership error.

What must endure is not Agreement Persistence, but Correction Regeneration: the capacity to regenerate structures of correction.

Long-term observation of the China–US AI incident channel should therefore not merely count the years it has existed. It should examine whether changing conditions leave it capable of redefining triggers, revising permissions, reaffirming shared risks, and re-establishing minimum communication.

V. Agreements constrain existing strategic games; self-interest first keeps generating new ones

Agreements, laws, deterrence, and institutions usually operate after purposes form: actors have decided to be safe, lead, or win, and rules then tell them which means are prohibited and which boundaries cannot be crossed. As long as self-interest-first purpose generation remains unchanged, agreements face an internal pressure: once circumventing the rules becomes more advantageous, why continue complying?

This does not make agreements useless. Provisional peace, crisis communication, arms control, and institutional constraints may be highly important. The problem is that they chiefly constrain games already generated, without automatically changing how the next generation of actors generates purposes anew.

Agreements can constrain strategic games already under way, but cannot prevent a self-interest-first logic from continually generating new ones. Changing the long-term trajectory of these games ultimately requires reaching the first-order logic that generates purposes.

“Internal” here is not a moral slogan. It means that shared consequences, others' circumstances, and conditions for future correction no longer enter only as external costs after purposes form. Before a purpose is locked in, they can change whether that purpose should continue forming at all. If new actors can repeatedly regenerate this way of generating purposes, civilization passes on not only fixed clauses but an underlying way of producing new clauses, cooperation, and boundaries.

VI. What REAL-ACT 01 now actually tests

This case does not undertake to prove an agreement can remain stable forever. It observes three clearer questions:

- First, can the AI/SI incident channel identify risks that neither side can afford to realize within the broader competition, bringing them into decisions before local wins and losses?
- Second, when actors, interests, power, or technology change, can the channel of correction reopen itself rather than merely demand future obedience to today's clauses?
- Third, if communication and correction exist only while both sides currently see them as advantageous, they remain chiefly a provisional equilibrium. Only if new actors under new conditions regenerate reasons to preserve an afterward do they approach the internally generated stability this research concerns.

All three questions allow failure. Failure does not automatically disprove a grand theory, but reveals whether the practical obstacle lies in purposes, incentives, permissions, technology, actor relationships, or continuity across time.

VII. Interpretive discipline and routes for participation

This case does not undertake to prove that no second path could possibly exist in the universe, still less to prove that a second path does exist. Different paths may be proposed; once genuinely proposed, they can enter discussion and testing.

The main line remains to distinguish real facts from inferences and place the causal chain—agreements constrain existing strategic games while self-interest first continually generates new ones—within a real-world window.

Participation may take the form of reconstruction, falsification, alternatives, or implementation. Its value is not in voting for a theory but in bringing more actors into the same real problem.

What matters is not more people believing the same answer, but more people genuinely thinking through the same question.

Mode of participation | Actual contribution

Reconstruction | Independently retrace the causal chain: why does time erode agreements, and what must be preserved over the long term?

Falsification | Identify a causal step that fails or a conflict between real evidence and the judgment.

Alternative | Propose a fundamentally different path; prior acceptance of this essay's terminology is not required.

Implementation | Turn specific solvable problems into real actions that can be verified, stopped, and reversed.

VIII. The next observation window: November is only a near-term point; the course of time is the test

Public Chinese and US materials indicate that the next AI/SI dialogue is planned for November 2026.[1][2] This is a near-term observation point, not the case's endpoint. In the short term, observe whether clearer incident categories, receiving actors, verification, false-alarm handling, or crisis-communication arrangements emerge. Over the longer term, observe whether these mechanisms endure deteriorating relations, replacement of actors, shifts in power, and technological upgrades.

Subsequent evidence status must therefore continue to distinguish political commitments, operating rules, actual use, short-term effects, persistence across crises, persistence across actors, and the capacity to regenerate correction. Establishing one layer cannot automatically stand in for the next.

IX. Ultimately, this case asks more than how long this agreement can live

REAL-ACT 01 has moved from asking whether China and the United States can establish an AI incident communication channel to asking whether civilization can regenerate correction before irreversibility once the actors, interests, power, and technological conditions on which an agreement rests have changed.

If the answer relies only on longer agreements, stronger penalties, and another round of external constraints, the old structure still uses downstream rules to repair results generated by upstream logic. Our direction continues to ask how reasons not to push one another into irreversibility can enter purpose generation earlier and remain regenerable after new actors appear.

Civilization Leap is not the establishment of an eternally valid agreement. It is enabling future people, organizations, states, and intelligences, even under conditions unlike today's, to regenerate relational structures that do not push one another into irreversibility.

Public sources and adjacent research

[1] The White House. "Fact Sheet: President Donald J. Trump Advances a Fair and Reciprocal Relationship with China While Hosting Historic State Visit." 25 Sep 2026. <https://www.whitehouse.gov/fact-sheets/2026/09/fact-sheet-president-donald-j-trump-advances-a-fair-and-reciprocal-relationship-with-china-while-hosting-historic-state-visit/>

[2] Ministry of Foreign Affairs of the People's Republic of China. "China and the United States Reach an Eight-Point Consensus on Outcomes." 2026-09-26. https://www.fmprc.gov.cn/zyxw/202609/t20260926_12031616.shtml

[3] Ministry of Foreign Affairs of the People's Republic of China. "President Xi Jinping Holds Talks with U.S. President Donald J. Trump." 25 Sep 2026. https://www.mfa.gov.cn/eng/xw/zyxw/202609/t20260925_12031181.html

[4] Fearon, J. D. "Rationalist Explanations for War." *International Organization* 49(3), 1995, 379–414. DOI: 10.1017/S0020818300033324.

[5] Leeds, B. A. "Alliance Reliability in Times of War: Explaining State Decisions to Violate Treaties." *International Organization* 57(4), 2003, 801–827. DOI: 10.1017/S0020818303574057.

[6] Powell, R. "War as a Commitment Problem." *International Organization* 60(1), 2006. The article analyzes how large and rapid shifts in the distribution of power can create commitment problems even when current bargains are preferable to war.

[7] Chernavskikh, V. & Palayer, J. Impact of Military Artificial Intelligence on Nuclear Escalation Risk. SIPRI Insights on Peace and Security, No. 2025/06, June 2025. DOI: 10.55163/FZIW8544.

Companion reading

- Real-world overview: “A Strike and the Whole Network It Moves,” in Seeing the Whole Picture: Sixteen Entries, One Network; Crisis Emergence: New Structural Forms of Interstate Strategic Interaction in the AGI Era, v4.0.
- Real-world governance and overall strategy: Bloc Formation in AI Governance and Humanity's Overall Security, v2.0; The Only Path: States' Long-Term Interests and Civilization Leap in the AI Era, v6.21.
- Underlying logic and action: Agreements Constrain Strategic Games; Self-Interest First Keeps Generating New Ones, V1.1; When Rules Cannot Catch Up, What Can We Do Now?, open guide V1.1.

Author: Zijun Fu | Produced by the Civilization Leap Research Group